Appendix A

Data Quality Indicators

Definitions and Calculations

Data Quality Indicators

Definitions and Calculations

The terms precision, accuracy, bias, comparability, representativeness, and completeness are all terms that represent qualitative concepts. Precision, accuracy, bias, and completeness are, for the purposes of the CASTNet project, associated with quantitative measurements.

Comparability and representativeness are assessed using indirect methods that provide weight of evidence via comparison with generally accepted standards.

Precision

The definition of precision is taken from International Organization of Standardization (ISO) 3534-1 "… the closeness of agreement between independent test results obtained under stipulated conditions." CASTNet uses measurements from collocated site pairs and duplicate analyses of laboratory samples to assess precision.

The **M**ean of the **A**bsolute value of single or aggregated **R**elative **P**ercent **D**ifference (MARPD) is used to express precision and is calculated as shown below:

$$MARPD = \frac{1}{k}\sum_{j=1}^{k}\left(\frac{|S1 - S2|}{S1 + S2}\right)_j x\ 200$$

Where:

S1	=	The value for the primary sampler,
S2	=	The value for the collocated sampler, and
k	=	The number of pairs of valid data.

For reporting purposes, the absolute value of the relative percent difference is used when a single pair is evaluated and referred to simply as ARPD or RPD. The formula shown above then reduces to:

$$RPD = \left(\frac{|S1 - S2|}{S1 + S2}\right) x\ 200$$

Note:
Signed results (positive and negative) are not generally used for reporting. An exception to this is in the reporting of bias as discussed later.

Accuracy

Accuracy is defined as the closeness of agreement between a "true" or reference value and an associated measurement result. CASTNet uses certified references traceable to NIST to obtain the true value used for assessment.

$$\%R = \left[1 + \left(\frac{Y - X}{X}\right)\right] x100$$

Where:

%R	=	Percent recovery. The amount measured as compared to the "true" value, expressed as a percentage,
Y	=	The measured value, and
X	=	The true value.

Bias

Bias is defined as a systematic error in measurement wherein the measured value displays a consistent positive or negative error as compared to a true value. Bias measurements are calculated either as a percent difference or as a mean arithmetic difference. The signed arithmetic difference is used for assessment where values are too small or too close to the limit of detection to calculate a meaningful percent difference.

$$\%D = \frac{1}{k}\sum_{j=1}^{k}\left(\frac{Y - X}{X}\right)_j x100$$

Where:

%D	=	Percent difference. The difference between the amount measured and the "true" value, expressed as a percentage,
Y	=	The measured value,
X	=	The true value, and
k	=	The number of valid comparisons.

And:

$$MAD = \frac{1}{k}\sum_{j=1}^{k} (Y - X)_j$$

Where:

MAD	=	Mean arithmetic difference. The arithmetic difference between the amount measured and the "true" value,
Y	=	The measured value,
X	=	The true value, and
k	=	The number of valid comparisons.

Completeness

Completeness is defined as the percentage of valid data points relative to total possible data points.

$$\%C = \left[1 + \left(\frac{Y - X}{X}\right)\right] x100$$

Where:

%C	=	Percent completeness,
Y	=	The number of valid data points, and
X	=	The total possible number of data points.

Comparability

EPA guidance document QA/G-5 defines comparability as a "qualitative term that expresses the confidence that two data sets can contribute to a common analysis and interpolation … whether two data sets can be considered equivalent in regard to the measurement of a specific variable or groups of variables." Comparability is established via the same methods used for ensuring representativeness plus the use of conventional and standard units for reporting. In addition, the CASTNet laboratory participates regularly in laboratory intercomparison studies wherein blind samples are supplied to a group of participating laboratories for analysis.

Representativeness

EPA guidance document QA/G-5 defines representativeness as "a measure of the degree to which data accurately and precisely represent a characteristic of a population parameter at a sampling point or for a process condition or environmental condition. Representativeness is a qualitative term that should be evaluated to determine whether *in situ* and other measurements are made and physical samples collected in such a manner that the resulting data appropriately reflect the media and phenomenon measured or studied." As noted in Chapter 2, CASTNet representativeness is established via adherence to specified siting criteria, uniformity in equipment procurement and deployment, and uniform implementation of SOP.

Note on References: Term definitions used in this appendix are quoted or adapted from EPA QA/G-5 and ISO 3534-1.

U.S. Environmental Protection Agency (EPA). 1998. Guidance for Quality Assurance Project Plans (QA/G-5), EPA/600/R-98/018, Office of Research and Development.

International Organization for Standardization (ISO). 1993. Statistics—Vocabulary and Symbols — Part 1: Probability and general statistical terms. International Organization for Standardization, Geneva, Switzerland, 1993.